Modeling the Dynamics of Chinese Spot Interest Rates☆

Yongmiao Hong [a,b] , Hai Lin [c,*], Shouyang Wang [d]

[a] Department of Economics and Department of Statistical Science, Cornell University

[b] Wang Yanan Institute for Studies in Economics, Xiamen University

[c] Department of Finance and Wang Yanan Institute for Studies in Economics, Xiamen University

[d] Academy of Mathematics and System Science and Center for Forecasting Science, Chinese Academy of Science

Abstract

Understanding the dynamics of spot interest rates is important for derivatives pricing, risk management, interest rate liberalization, and macroeconomic control. Based on a daily data of Chinese 7-day repo rates from July 22, 1996 to August 26, 2004, we estimate and test a variety of popular spot rate models, including single factor diffusion, GARCH, Markov regime switching and jump diffusion models, to examine how well they can capture the dynamics of the Chinese spot rates and whether the dynamics of the Chinese spot rates has similar features to that of the U.S. spot rates. A robust M-estimation method and a robust Hellinger metric-based specification test are used to alleviate the impact of frequent extreme observations in the Chinese interest rate data, which are mainly due to IPO. We document that GARCH, regime switching and jump diffusion models can capture some important features of the dynamics of the Chinese spot rates, but all models under study are overwhelmingly rejected. We further explore possible sources of model misspecification using some diagnostic tests. This provides useful information for future improvement on modeling the dynamics of the Chinese spot rates.

JEL Classification: E4, C5, G1.

Key Words: Generalized residuals, Robust specification tests, Robust M-estimation, Spot rate models

---

☆ We thank the editor, Richard T. Baillie, two anonymous referees, Max Chen and Zhenlong Zheng for their comments and suggestions. Hong and Wang acknowledge the support by the Cheung Kong Scholarship of the Chinese Ministry of Education and the Chinese National Science Foundation.

* Corresponding author. Hai Lin, Department of Finance and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, 361005, China. Email: cfc@xmu.edu.cn.

1. Introduction

The term structure of interest rates, which characterizes the relationship between yields on a zero coupon bond and time to maturities, plays a fundamental role in economics and finance, especially in macroeconomic policy making, derivatives pricing, hedging, and risk management for fixed income securities. The spot rate is the yield on a zero coupon bond with zero maturity and is the most important factor of the term structure of interest rate. It is important to understand the dynamics of spot rates over time. For example, the knowledge of the dynamics of spot rates is needed when we calculate the expected discounted value of uncertain future payoffs in pricing contingent claims. A vast literature has been devoted to modeling the dynamics of spot rates in mature markets. These include, among many others, Chan, Karolyi, Longstaff and Sanders (CKLS, 1992), Ait-Sahalia (1996, 1999), Gray (1996), Stanton (1997), Brenner, Harjes and Kroner (1996), Andersen and Lund (1997), Ahn and Gao (1999), Conley, Hansen, Luttmer and Scheinkman (1997), Chapman and Pearson (2000), Balduzzi and Eom (2000), Dai and Singleton (2000), Durham (2003), Durham and Gallant (2002), Ang and Bekaert (2002), Elerian, Chib and Shephard (2001), Das (2002), Jones (2003), Johannes (2004), Hong, Li and Zhao (2004) and Hong and Li (2005). These studies document some important features of spot interest rates in mature markets, particularly the U.S. markets. For example, there exists significant mean reverting when using one factor diffusion models for the U.S. interest rates, although whether there exists a nonlinear drift is inconclusive. Ait-Sahalia (1996), Stanton (1997), Conley et al. (1997), Ahn and Gao (1999) report evidence of nonlinear drifts, whereas Chapman and Pearson (2000), Pritsker (1998), Hong et al. (2004) cast some doubts on it. Chan et al. (1992) and Hong et al. (2004) document that the interest rate volatility tends to be higher when the interest rate level is higher, which is called “level effect” and often characterized by a Constant Elasticity Variance (CEV) specification. Moreover, Brenner et al. (1996) and Andersen and Lund (1997) find that it is important to capture conditional heteroscedasticity of interest rates via stochastic volatility/GARCH models, which outperform one factor spot rate models. On the other hand, Gray (1996), Ang and Bekaert (2002), Das (2002) and Johannes (2004) find that that regime switching and jump models help capture volatility clustering and especially the excess kurtosis and heavy tails of spot interest rates. Once stochastic volatility/GARCH, regime switching, or jump effects are introduced, the importance of modeling mean reversion in drift diminishes substantially. Sophisticated specification for the drift usually has little impact on overall goodness of fit of spot rate models (Durham 2003).

While the spot rate dynamics has been well examined in the mature markets like U.S. markets, there has been little study on spot interest rates in China and other emerging markets. To our knowledge, there has been no pioneering work on modeling the dynamics of the Chinese spot rates. This is perhaps due to the relatively short history of the Chinese bond markets, and the strict government regulation on the Chinese interest rates. The main purpose of this paper is to characterize the dynamics of the Chinese spot rates. In particular, we are interested in whether the Chinese spot rates share similar dynamic features to the U.S. spot rates, and whether the models which can capture important features of the U.S. interest rate dynamics can also characterize important features of the Chinese spot rates.

With the continuing economic reforms over the past 30 years and the recent entry of WTO, Chinese economy is becoming more and more market-oriented, including interest rate liberalization. Understanding the dynamics of the Chinese spot rates is important for developing efficient financial markets, determining effective interest rate policy and piloting optimal investments over different time horizons. Moreover, knowledge of the Chinese interest rate dynamics aids in the determination of security prices, prediction of interest rate changes and choice of hedging strategies. Generally speaking, in such an emerging market as China, the spot rate plays a role similar to the FED fund rate in the U.S., and it is a fundamental instrument in developing bond markets and other fixed income security markets.

In this paper, we provide a first comprehensive empirical study on the dynamics of the Chinese spot rates. We consider a wide variety of spot rate models, including single factor diffusion, GARCH, Markov regime switching and jump diffusion models, and examine how well they can capture important features of the Chinese spot rates. To reduce the impact of frequent extreme observations in the Chinese interest rates mainly due to IPO, we use a robust M-estimation method. Similarly, we use a robust nonparametric test proposed by Hong and Li (2005) and Hong, Li and Zhao (2007) to test the adequacy of these models for the Chinese spot rates. We find that there exists significant mean-reverting in the Chinese spot rates, with a noticeable nonlinear drift. There also exists significant volatility clustering which can be captured by a level effect model or a GARCH model, but it does not help much when combining both the level effect and GARCH effect together. It is also documented that regime switching and jump models can help capture volatility clustering and particularly the frequent extreme observations. Nevertheless, all models under study are firmly rejected.

In Section 2, we review the history of the Chinese interest rate liberalization and describe the data on the Chinese spot rates. In Section 3, we introduce a wide variety of spot rate models and a robust M-estimation method. In Section 4, we describe the robust specification tests by Hong and Li (2005) and Hong et al. (2007). In Section 5, we describe the goodness of fit of each model. We subject each model to specification evaluation and diagnostic check in Section 6. Section 7 concludes the paper.

2. Interest rate liberalization in China and proxy for the Chinese spot rates

2.1 Interest rate liberalization in China

China regulated saving rates with different maturities until mid-1980s. Since the set up of the stock market and bond market in late 1980s and 1990s, the interest rate has gradually become an important instrument in macroeconomic control, risk management and asset pricing. However, due to the short history of the Chinese market economy and the main focus on developing the stock market, the Chinese bond market and interest rate liberalization are underdeveloped. The spot interest rate in China is determined in two main markets, i.e., the inter-bank borrowing market and bond repurchase market. Chinese inter-bank borrowing markets appeared in 1980s at different locations over China and were united into a single market in January, 1996. On March 1, 1996, the Chinese government set up a framework of two-level inter-bank borrowing market. The first level consists of the headquarters of 15 commercial banks and 35 financing centers, while the second level includes bank branches and other financial organizations. The uniform borrowing rates in this market are named as "CHIBOR". The upper limit of CHIBOR was removed in 1996 so that it could reflect the information of financial market more closely. CHIBOR mainly consists of short term interest rates, with 4 months as the longest maturity. In 2000, the 1-day and 7-day inter-bank borrowings accounted for 71.4% of the total inter-bank borrowing. Therefore, CHIBOR mainly characterizes the Chinese short term interest rates.

Chinese bond repurchase began in 1991 at four stock exchanges, i.e., Shanghai Stock Exchange, Wuhan Stock Trading Center, Tianjin Stock Trading Center, and the STAQ system (the later three were closed later). In 1997, to prevent banks from investing in stock markets, the Chinese central bank—the People's Bank of China prohibited all commercial banks from the bond repurchase on stock exchanges and opened another bond repurchase sub-market in the inter-bank market. This leads to two independent and segmented bond repurchase markets in China, i.e., the OTC market at inter-bank markets and the electronic market at stock exchanges. These bond markets are artificially segmented, with different interest rates for the same bond.

The institutional members engaging in the inter-bank repurchase are far more than those in the inter-bank borrowing.[1] Moreover, the repurchase is mortgaged borrowing, with credit risk less than credit borrowing. As a result, the bond repurchase market is more active. Since 1999, the trading volume of repurchase has been much higher than that of inter-bank borrowing, as shown in Table 2. Moreover, the interest rate there is more stable, making it more representative as the Chinese spot interest rate.

The long term interest rates are determined by the Chinese long term bond market. Like the spot interest rate markets, there are two segmented long term bond markets, the OTC bond market at the inter-bank market and the electronic market at stock exchanges. However, interest rates of middle maturities are controlled tightly by the Chinese central bank. They do not change every day to reflect the market information and remain unchanged for a relatively long period. They change only when the Chinese government uses them as macroeconomic instruments.

There are two main deficiencies of the current interest rate mechanism in China that hinder the play of its fundamental roles in the Chinese economy. First, there exist two independent bond

[1] Up to 2007, there are 274 members in the Chinese inter-bank borrowing market, while there are more than 800 members in the Chinese inter-bank repurchase market.

markets that share similar functions and trade same products, i.e., the inter-bank OTC market and the exchange electronic market. Since they are artificially segmented, a same bond has different prices at these two markets, resulting in two different interest rates between the inter-bank market and the exchange market. The difference in the interest rate levels of two segmented markets reflects different expectations of investors. It is very difficult, if not possible, to develop derivative markets without a uniform market interest rate.

Second, the deposit rates in China are still regulated by the Chinese central bank. They cannot be changed by commercial banks to reflect market information. Therefore, there is a large gap between the regulated deposit rates and the market interest rates, and serious problems and arbitrage opportunities may arise. For example, if the deposit rate is lower than the market rate for the bond with same maturity, some large investors would borrow money from the banks to invest on the bond market and construct an arbitrage portfolio.

The Chinese government has recently proposed several reforms on interest rate liberalization. It issues bonds at both the inter-bank market and the exchange market. Some security companies and trust companies are allowed to enter the inter-bank market to join the issuing. The Chinese central bank also introduced the market maker system in 2001 on secondary markets, allowing some eligible banks to be the bid-ask quoters that have a similar function to market makers.

To construct a well-functioning interest rate term structure, the Chinese government begins to issue bonds ranging from long terms to short terms. By issuing and trading bonds with different maturities, an integrated bond market can be developed to provide a robust benchmark for pricing and hedging. Furthermore, the Chinese government has the plan to gradually deregulate deposit rates and liberalize them eventually. It also tries to introduce other financial instruments, such as Bond Futures, Stock Index Futures and Monetary Market Fund (MMF). In all, although the Chinese interest rate liberalization is still far from complete, it has been advancing steadily. Table 1 summarizes the major characteristics of the Chinese interest rate liberalization including its histories and recent developments.

2.2 Proxy for the Chinese spot rates

To investigate the dynamics of the Chinese spot rates, we shall use the 7-day repo rates in the Chinese exchange market as the proxy of the Chinese spot rates.[2] Table 2 reports the trading volumes of 1-day repo, 7-day repo, 14-day repo and 1-month repo that could be the representative candidates of the Chinese spot rate during the sample period. It also reports the trading volume of the Chinese inter-bank borrowing market. These data are obtained from the WIND dataset and the *Chinese Financial Industry Annual Report*. The trading of the repo market is much more active than that of the inter-bank borrowing market in most years except 1999. The trading of 7-day repo

[2] In empirical studies of spot rate models in mature markets, yields on different short term debts are used as proxies of spot rates. These include 1-month T-bill rates used by Gray (1996) and Chan et al. (1992) and Hong et al. (2004), 3-month T-bill rates used by Stanton (1997) and Andersen and Lund (1997), 7-day Eurodollar rates used by Ait-Sahalia (1996) and Hong and Li (2005), and the Fed fund rates used by Conley et al. (1997) and Das (2002).

is the most active among all repos, which makes it as the best proxy of the Chinese spot interest rates. The transaction of 7-day repo in the inter-bank market began only from 1999. Moreover, the inter-bank market is an OTC market and the quoted price may not reflect the actual transaction price due to private negotiations between traders. The number of participants in the OTC market is also smaller than that in the exchange market. We use the daily data of 7-day repo rates from July 22, 1996 to August 26, 2004 in Shanghai stock exchange, with a total of 1954 observations. Because of the influence of holidays on the repurchase time, the original data do not exactly represent the 7-day repo rates. For instance, one 7-day repurchase buyer will generally repurchase the bond at a prespecified price in 7 days. However, if in 7 days the market is closed due to holidays or other reasons, the repurchase is delayed to the next working day, while the repurchase price and total interest remain unchanged. Thus, the investor could use the fund for more than 7 days while only paying the 7-day interest. Since this information is public, the 7-day repo rate will increase to counteract the delay of repurchase and interest payout. To eliminate such effect, we transform the original data:

$$r_t = \frac{\tilde{r}_t \times 7}{\tau} \tag{2.1}$$

where r_t is the exact 7-day repo rate after transformation, $\tilde{r}_t$ is the listed 7-day repo rate, and τ is the number of exact repurchase days.

Figure 1 plots the level and change series of the transformed daily 7-day repo rates, as well as their histograms. There is persistent volatility clustering, and in general, the volatility was higher at the higher interest rate level before 1999, i.e., there exists the “level effect”. There appeared a change on the repo rate behavior after 1999. There may be several reasons for this structure break. During 1996 to middle 1999, the central bank decreased the regulated saving rates 6 times. The 1-year saving rates declined from more than 7% to about 2% during this period. However, since then, the central bank has changed the saving rates much less frequently (only once), and the saving rates kept stable at about 2%.[3] The interventions of the Chinese central bank undoubtedly had a significant impact on the Chinese spot rates. Before 1999, the Chinese IPO price was determined by a rule that the IPO price was not higher than 15 times of earnings per share. The Chinese Securities Law exercised on July 1, 1999, however, reformed the IPO pricing mechanism, requiring that the IPO price should be based on the market value. This reform had a significant impact on the repo market. Moreover, another segmented market, the inter-bank market, also began to trade 7-day repos in 1999.

The marginal distribution of the interest rate level is skewed to the right, with a long right tail. The minimum and maximum interest rate levels during the sample period are 0.087% and

[3] The changes of 1-year regulated saving rates during the sample period are as follows: August 23, 1996: 7.47%; October 23, 1997: 5.67%; March 25, 1998: 5.22%; July 01, 1998: 4.77%; December 07, 1998: 3.78%; June 10, 1999: 2.25%; February 21, 2002: 1.98%.

30.00% respectively. The daily changes of repo rates also exhibit a high peak around 0. The repo rates show frequent jump behaviors, which are quite different from mature markets where the interest rates change stably most of the time. This is mainly due to the arbitrage behavior of large institutions in the Chinese IPO. Because of serious underpricing of IPO stocks on the primary market, the price of new issued stocks may increase more than 100% on the first listed day on the secondary markets. Before 1999, the return from bidding IPO stocks on the primary market and selling it immediately on the secondary markets could be as high as 100%. Then when there was an IPO on the primary market, the investors would demand a large amount of money for a few days at a rate as high as 30%, which results in a sudden jump of the repo rate. After IPO, the spot rate fell immediately. Other possible reasons for the extreme interest rate observations include liquidity shocks and interventions of the Chinese central bank. The repo market is less liquid than the stock market and may be subject to some liquidity problem on a particular day, which will result in large change of repo rates. The intervention of the Chinese central bank, such as operations in open markets and changes of reserve rates, may also affect the repo market in a sudden way.

Nevertheless, IPO is the main reason for frequent large changes in the Chinese spot rates. Figure 2 plots the dynamics of the Chinese 7-day repo rates with IPO during the sample period. Jumps of interest rates happened frequently on IPO days, especially before 2002. After 2002, such phenomena disappeared because the degree of IPO underpricing decreased gradually. This suggests a structure break of IPO impacts on the Chinese spot rates before and after 2002. The mean 7-day repo rate for the days with IPO is 5.91% and the mean 7-day repo rate for the days without IPO is only 3.92%. The difference is 1.99%, which is significant at the 1% level. This suggests that IPO does affect the Chinese 7-day repo rates.

3. Spot rate models

We will examine whether some popular dynamic models that have been used to capture the dynamics of spot rates in mature markets can also be used to characterize the Chinese spot rates. The models to be examined include single factor diffusion, GARCH, regime switching, and jump diffusion models. We now introduce these models and a robust M-estimation method for them.

3.1 Single factor diffusion models

One popular class of spot rate models is single factor diffusion models, which have been widely used in modern finance and fixed-income securities pricing. For some single factor diffusion models, such as the Vasicek model and CIR model, the prices of discounted bonds have a closed form expression, which offers a lot of convenience in pricing other interest rate derivatives.

Specifically, the spot rate is assumed to follow a single factor diffusion,

$$dr_t = \mu(r_t,\theta)dt + \sigma(r_t,\theta)dW_t, \tag{3.1}$$

where $\mu(r_t,\theta)$ and $\sigma(r_t,\theta)$ are the drift and diffusion functions, W_t is a standard Brownian motion. Here, $\mu(r_t,\theta)$ and $\sigma(r_t,\theta)$ completely determine the model transition density, which captures the full dynamics of r_t.

We consider a variety of discretized single factor diffusion models which are nested by Ait-Sahalia (1996) nonlinear drift model,

$$\begin{cases}\Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t, \\ \{z_t\} \sim iid.N(0,1),\end{cases} \tag{3.2}$$

where $\Delta r_t = r_t - r_{t-1}$. Discretizations are approximations of continuous time models. Nevertheless, Bandi (2002) documents that the error introduced by discretizing is of second-order importance if changes are measured over very short periods of time. Stanton (1997) and Das (2002) also document that the discretization bias for daily data we shall use is not substantial. To examine different model specifications, we allow the drift function to have a zero, linear, and nonlinear specification respectively and allow the diffusion function to be a constant or depend on the interest rate level, which is referred to as the "level effect". The diffusion specification σr_{t-1}^{ρ} is called the Constant Elasticity Variance (CEV). For convenience, all single factor diffusion models examined are listed in Table 3(a).

3.2 GARCH models

Despite the popularity of single factor diffusion models, many studies (e.g., Brenner et al. 1996; Andersen and Lund 1997) have documented that single factor diffusion models fail to capture the well-known persistent volatility clustering of financial returns including interest rates. Brenner et al. (1996) examine various GARCH models for the U.S. interest rates and find that GARCH models significantly outperform single factor diffusion models.

To evaluate the importance of GARCH specifications in modeling the Chinese spot rates, we consider six GARCH models listed in Table 3(b), including three drift specifications (zero, linear and nonlinear) and two volatility specifications (pure GARCH and combined CEV-GARCH). These models are nested by the following specification:

$$\begin{cases}\Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t}\, z_t, \\ h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-2}^{2\rho} z_{t-1}^2), \\ \{z_t\} \sim iid.N(0,1).\end{cases} \tag{3.3}$$

Various GARCH models allow us to examine the contribution of the drift term in modeling the Chinese spot rates in the presence of GARCH or CEV, and to examine the additional contribution of the GARCH effect beyond the CEV effect. For identification, we set $\sigma = 1$ in all GARCH models.

3.3 Markov regime switching models

Due to the change of monetary policy, business cycle and other macroeconomic conditions, the dynamics of interest rates may change over time. Based on this motivation, Bansal and Zhou (2002), Gray (1996), Ball and Torous (1998), Ang and Bekaert (2002), and Sanders and Unal (1988) use Markov regime switching models for the U.S. interest rates. Like these studies, we examine a class of two-regime models for the Chinese spot rates, where the latent state variable s_t follows a two state, first order Markov chain. We refer to the regime in which s_t =1 (or 2) as the first (or second) regime. Following Ang and Bekaert (2002), the transition probability of $\{s_t\}$ is assumed to depend on the one-lagged spot rate level,

$$\Pr(s_t = l \mid s_{t-1} = l) = \frac{1}{1+\exp(-c_l - d_l r_{t-1})}, \quad l = 1,2 \tag{3.4}$$

Table 3(c) lists a variety of regime switching models, all of which are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1}(s_t) r_{t-1}^{-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho(s_t)} \sqrt{h_t}\, z_t, \\ h_t = \beta_0 + \beta_1 E[e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2})]^2 + \beta_2 h_{t-1}, \\ e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2}) = [\Delta r_{t-1} - E(\Delta r_{t-1} \mid r_{t-2}, s_{t-1})] / \sigma(s_{t-1}), \\ \{z_t\} \sim iid.N(0,1), \end{cases} \tag{3.5}$$

We consider three specifications for the drift function: zero, linear and nonlinear drifts respectively, and three specifications of the diffusion function: CEV, GARCH and CEV-GARCH, respectively. Thus, we have a total number of nine regime switching models. Different from Gray (1996), we use the same GARCH specification across different regimes. While many previous studies using GARCH models set the elasticity parameter to 0.5 for U.S. interest rate data, we allow it to be regime-dependent and estimate it from data. Similarly, for identification, we set the diffusion constant $\sigma(s_t)$ =1 for s_t =1.

It can be shown that the conditional likelihood of the interest rate r_t in a regime switching model is

$$p(\Delta r_t \mid I_{t-1}) = \sum_{l=1}^{2} p(\Delta r_t \mid s_t = l, I_{t-1})\, p(s_t = l \mid I_{t-1}) \tag{3.6}$$

where $p(s_t = l|I_{t-1})$, the *ex ante* probability that the data are generated from regime l at time t, can be computed using Bayes' rule via a recursive procedure (Hamilton 1989). Therefore, the conditional distribution of a regime switching model is a mixture of two normal distributions, which offers great flexibility in modeling skewness, kurtosis and heavy tails.

3.4 Jump diffusion models

Various economic shocks, IPO, news announcement, government policy changes, and the interventions of central banks on financial markets, may affect the spot rates in a sudden way and generate interest rate jumps. Baz and Das (1996) discuss the estimation of jump diffusion models by maximum likelihood method (MLE). Das (2002) and Johannes (2004) show that diffusion models with stochastic volatility cannot explain the excessive leptokurtosis exhibited in the changes of U.S. spot rates, but jump diffusion models can capture such features.

We consider a class of discretized jump diffusion models listed in Table 3(d). We consider zero, linear and nonlinear drift specifications respectively. For volatility, we consider CEV, GARCH and combined CEV-GARCH specifications respectively. These nine models are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t + J(\psi, \gamma^2)\pi(q_t), \\ h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1} \mid r_{t-2})]^2 + \beta_2 h_{t-1}, \\ \{z_t\} \sim iid.N(0,1), \\ \{\pi(q_t)\} \sim iid.Bernoulli(q_t), \\ J \sim N(\psi, \gamma^2), \end{cases} \tag{3.7}$$

where J is a random jump size and q_t is the jump probability with

$$q_t = \frac{1}{1 + \exp(-c - dr_{t-1})}. \tag{3.8}$$

Similar to regime switching models, the conditional distribution of a jump diffusion model is also a mixture of two normal distributions. However, the specifications of regime switching models are more sophisticated. In (3.5), all drift parameters are regime dependent, whereas in (3.7) only the intercepts in the conditional mean and variance are different. For identification, we set $\sigma = 1$ in all GARCH and CEV-GARCH specifications.

3.5 Robust M-estimation

The existence of outliers may substantially affect model parameter estimation. Dell'Aquila, Ronchetti and Trojani (2003) and Czellar, Karolyi and Ronchetti (2007) propose robust estimations in a GMM framework. The GMM approach may be quite difficult for estimating some sophisticated models considered here, such as regime switching and jump diffusion models. Instead, we use a MLE that is robust to outliers. This is a robust M-estimator due to Huber (1981). Rather than assuming an i.i.d. normal distribution for the stochastic error term, it assumes that the error distribution is Gaussian for small values of the error and Laplacian for larger values of the error. Specifically, Huber (1981) proposes a robust likelihood function:

$$f(\varepsilon) = \begin{cases} \dfrac{\beta}{\sigma} \exp\left(-\dfrac{\varepsilon^2}{2\sigma^2} \right), & \text{for } |\varepsilon/\sigma| \le a \\ \dfrac{\beta}{\sigma} \exp\left(-\dfrac{a}{\sigma}\left(|\varepsilon| - \dfrac{a\sigma}{2} \right) \right), & \text{for } |\varepsilon/\sigma| > a \end{cases} \tag{3.9}$$

where a is the robustness parameter, which usually take values between 1 and 3, σ is the scale parameter, and β is the normalizing constant,

$$\beta = \frac{1}{2\left[\frac{\exp\left(-a^2/2\right)}{a} + \int_0^a \exp\left(-\frac{z^2}{2}\right)dz\right]} \tag{3.10}$$

Obviously, when $|\varepsilon/\sigma|>a$, $-\frac{a}{\sigma}\left(|\varepsilon|-\frac{a\sigma}{2}\right) > -\frac{\varepsilon^2}{2\sigma^2}$. Therefore, the robust MLE reduces the impact of outliers on parameter estimation by a truncated density function. This is in essence similar to the approaches in Dell'Aquila, Ronchetti and Trojani (2003) and Czellar, Karolyi and Ronchetti (2007). In our application, we set a = 2.

To account for the structure break in 1999, we introduce dummy variables of the drift, volatility, elasticity and jump probability parameters, i.e., let α_D, σ_D, ρ_D, c_D and d_D be 1 before 1999 and zero after 1999.

4. Nonparametric evaluation method

To evaluate the relative performance of spot rate models, we use a robust nonparametric test proposed by Hong and Li (2005). Suppose $\{r_t\}$ has an unknown true conditional probability density function $p_0(r|I_{t-1})$, where I_{t-1} is the information set available at time t-1. For all spot rates models introduced above, there exists a model-implied conditional density function $p(r|I_{t-1},\theta)$. If a dynamic model adequately characterize the full dynamics of $\{r_t\}$, the conditional density model $p(r|I_{t-1},\theta)$ will coincide with the true conditional density $p_0(r|I_{t-1})$ for some unknown true parameter value θ_0. Thus, one can assess the adequacy of a spot rate model by measuring the distance between $p(r|I_{t-1},\theta)$ and $p_0(r|I_{t-1})$. Suppose we have a random sample $\{r_t\}_{t=1}^n$ of size n, Hong and Li (2005) consider the following probability integral transform

$$Z_t(\theta) \equiv \int_{-\infty}^{r_t} p(r,t\,|I_{t-1},\theta)dr,\ t = 1,2,...,n\,. \tag{4.1}$$

The transformed series $\{Z_t \equiv Z_t(\theta_0)\}_{t=1}^n$ is i.i.d. U[0,1] under correct model specification. The sequence $\{Z_t(\theta)\}_{\tau=1}^n$ is often called the "generalized residuals" of the model $p(r|I_{t-1})$. Intuitively, the i.i.d. U[0,1] property captures two important aspects of model specification: i.i.d. characterizes the correct specification of model dynamics, and U[0,1] characterizes correct specification of the model stationary distribution.

To test i.i.d. U[0,1], Hong and Li (2005) develop two nonparametric tests of i.i.d. U[0,1] by comparing a kernel joint density estimator $\hat{g}_j(z_1,z_2)$ of $\{Z_t, Z_{t-j}\}$ with unity, the product of two U[0,1] densities. The kernel joint density estimator $\hat{g}_j(z_1,z_2)$ for any integer $j>0$ is defined as follows:

$$\hat{g}_j(z_1,z_2)=(n-j)^{-1}\sum_{t=j+1}^{n}K_h(z_1,\hat{Z}_t)K_h(z_2,\hat{Z}_{t-j}), \tag{4.2}$$

where

$$K_h(x,y)=\begin{cases} h^{-1}k(\frac{x-y}{h})/\int_{-(x/h)}^{1}k(u)du, x\in[0,h), \\ h^{-1}k(\frac{x-y}{h}), x\in[h,1-h], \\ h^{-1}k(\frac{x-y}{h})/\int_{-1}^{(1-x)/h}k(u)du, x\in(1-h,1], \end{cases}$$

and the kernel $k(.)$ is a bounded symmetric probability density with support [-1,1]. One example is the quadratic kernel $k(u)=(15/16)(1-u^2)\mathbf{1}(|u|\le 1)$, where $\mathbf{1}(|u|\le 1)$ is the indicator function, taking value 1 if $|u|\le 1$ and value 0 otherwise. Also, $\hat{Z}_t=Z_t(\hat{\theta})$, and $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator for θ_0. Following Scott (1992), we choose $h=\hat{S}_Z n^{-1/6}$, where $\hat{S}_Z$ is the sample standard deviation of $\{\hat{Z}_t\}_{t=1}^n$.

The first test is a properly standardized version of the quadratic form between $\hat{g}_j(z_1,z_2)$ and 1, the product of two U[0,1] densities:

$$\hat{Q}(j)\equiv\frac{(n-j)h\int_0^1\int_0^1[\hat{g}_j(z_1,z_2)-1]^2dz_1dz_2-A_h^0}{\sqrt{V_0}}, \tag{4.3}$$

for some nonstochastic centering and scale factors A_h^0 and V_0 (see Hong and Li 2005).

The quadratic form test $\hat{Q}(j)$ might be adversely affected by any imprecise estimate for $\hat{g}_j(z_1,z_2)$ in finite samples, which may be caused by imprecise estimation of $\hat{\theta}$, data sparsity, and outliers. Compared to the U.S. spot rates, the Chinese spot rates have much more extreme observations. In order to alleviate the possible impact of extreme observations on model validation, we can use a test based on the squared Hellinger metric, which is a quadratic form between $\sqrt{\hat{g}_j(z_1,z_2)}$ and $\sqrt{1\bullet 1}=1$. The test statistic is given by

$$\hat{H}(j)\equiv\frac{4(n-j)h\int_0^1\int_0^1[\sqrt{\hat{g}_j(z_1,z_2)}-1]^2dz_1dz_2-A_h^0}{\sqrt{V_0}}. \tag{4.4}$$

This test is also proposed in Hong and Li (2005). Under correct model specification, both $\hat{Q}(j)$

and $\hat{H}(j)$ converges to N(0,1) in distribution as $n \to \infty$. Under model misspecification, they diverge to infinity with probability approaches one as $n \to \infty$ whenever $\{Z_t, Z_{t-j}\}$ are not independent or U[0,1]. Since $\hat{H}(j)$ uses the square root of $\hat{g}_j(z_1, z_2)$, it is expected to be more robust to outliers or sparseness in data. We use it in our study.

The use of $\hat{H}(j)$ with different *j's* can reveal the information on the lag orders at which there is a significant departure form i.i.d. U[0,1]. However, when comparing two different models, it is more desirable to construct a single portmanteau test. Otherwise we would run into difficulty when one model has smaller $\hat{H}(j_1)$ at lag j_1 but the other model has a smaller $\hat{H}(j_2)$ at lag $j_2 \neq j_1$. To avoid this, we can use a portmanteau test statistic:

$$\hat{W}(p) = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \hat{H}(j) \tag{4.5}$$

Following Hong et al. (2007), we can show $\hat{W}(p) \to N(0,1)$ in distribution under correct model specification. Under model misspecification, $\hat{W}(p) \to \infty$ in probability where there exists some order $j \in \{1, 2, \ldots, p\}$ such that Z_t and Z_{t-j} are not independent or U[0,1].[4]

The model generalized residuals $\{\hat{Z}_t\}$ contain rich information on potential sources of model misspecification and can be used for diagnostic analysis. The U[0,1] property measures how well a model captures the marginal distribution, while the i.i.d. property measures how well a model captures the dynamics. Hong and Li (2005) and Hong et al. (2007) consider the following test statistic:

$$\mathrm{M}(m,l) \equiv \left[\sum_{j=1}^{n-1} w^2(j/p)(n-j)\hat{\rho}_{ml}^2(j) - \sum_{j=1}^{n-1} w^2(j/p) \right] \Big/ \left[2\sum_{j=1}^{n-2} w^4(j/p) \right], \tag{4.6}$$

where $\hat{\rho}_{ml}(j)$ is the sample cross-correlation function between $\hat{Z}_t^m$ and $\hat{Z}_{t-|j|}^l$, and *w*(.) is a weighting function for lag order *j*.[5] An example of *w*(.) is the Bartlett kernel $w(z) = (1-|z|)\mathbf{1}(|z| \leq 1)$. Hong (1996), Hong and Li (2005) and Hong et al. (2007) show that for each given pair of positive integers *(m,l)*, M*(m,l)* → N(0,1) in distribution under correct

---

[4] Hong et al. (2007, Theorem 2) use the portmanteau test statistic based on $\hat{Q}(j)$. However, their asymptotic theory is also applicable to the portmanteau test statistic based on $\hat{H}(j)$.

[5] A factor of 2 in the denominator of (4.6) is missing in Hong et al. (2007).

model specification, provided that the lag truncation order $p \equiv p(n) \to \infty, \quad p/n \to \infty$. Although the moments of the generalized residuals are not exactly the same as those of the original process $\{r_t\}$, they are high correlated. In particular, the choice of *(m,l)* = (1,1), (2,2), (3,3), (4,4) is sensitive to autocorrelations in level, volatility, skewness and kurtosis of the original process $\{r_t\}$, respectively. Furthermore, the choice of *(m,l)* = (1,2) and (2,1) is sensitive to ARCH-in-mean and leverage effects. Different choices of order *(m,j)* can examine various dynamic aspects of the underlying process. Similar to $\hat{H}(j)$ and $\hat{W}(p)$, upper-tailed N(0,1) critical values are suitable for M*(m,l)*.

5. Model estimation

5.1 Robust M-estimation

We now use the robust M-estimation method described in Section 3.5 to estimate various spot rate models. Table 4 reports robust parameter estimates with estimated robust standard errors and robust log-likelihood values for discretized single factor diffusion models. The estimates of the drift parameters in Vasicek, CIR and CKLS models all show significant mean-reversion, with an estimated long run mean around 2.56% (estimate of $-\alpha_0/\alpha_1$). For other models such as random walk and nonlinear drift models, some drift parameters are not significant. For the Dothan model, the parameters are significant but the robust log-likelihood is the smallest. This is consistent with the estimation result for the U.S. spot rates (Hong et al. 2004). The contribution of a nonlinear drift is evident. The robust log-likelihood increases from 6066.62 to 6210.99 by introducing a linear drift in the pure CEV, and increases to 6320.99 if we use a nonlinear drift. This differs from Hong et al. (2004) who find that the additional contribution of a nonlinear drift over a linear drift is small for the U.S. interest rates. There is also evidence of level effect: all elasticity parameter estimates are significant. However, unlike some previous studies (e.g., CKLS 1992), which estimate the elasticity parameter to be about 1.5 for the U.S. interest rates, our elasticity parameter estimate is about 0.5 for the Chinese spot rates, which is consistent with the CIR model. The estimates of dummy variable coefficients between 1996 and 1998 suggest that both drift and volatility behave quite differently before and after 1999. The drift dummy coefficient α_D is significantly positive for Vasicek, CIR, CKLS and Ait-Sahalia's nonlinear drift models, implying a higher interest rate level before 1999. The volatility dummy coefficient σ_D is significantly positive for Vasicek and CIR models, while the level effect elasticity dummy coefficient ρ_D is significantly negative for pure CEV, CKLS and nonlinear drift models. Thus, the volatility between 1996 and 1998 is significantly higher, while the sensitivity of interest rate volatility on the interest rate level becomes stronger after 1999. There may be two reasons for such findings. First, borrowing and lending of the short term money was mainly through inter-banks before 1999. After 1999, the repurchase market replaces the inter-bank market as the dominant market of short term financing for large institution investors. The short term financing of such large institution investors is more influenced by the level of the market interest rate. As a result, the sensitivity of the interest rate change to the interest rate level becomes stronger. On the other hand, the Chinese Security Regulation Commission (CSRC) reforms the IPO mechanism and imposes strict regulations on the flowing of bank money into the stock market after 1999. As a result, the degree of IPO under-pricing decreases gradually, which reduces the demand of a large amount of money for arbitrages. The interest rate volatility and jump probability become smaller.

Table 5 reports the estimation results of GARCH models, which outperform singe factor diffusion models. The robust log-likelihood increases from less than 6400 to more than 6500. All GARCH parameter estimates are significant. The sum of two GARCH parameter estimates, $\beta_1 + \beta_2$, is slightly larger than 1 when the level effect is not considered. With the level effect, $\beta_1 + \beta_2$ increases to some extent. However, it is possible that the spot rate model remains strictly stationary (Nelson 1991). The level effect in the GARCH model is significant with an estimate of about 0.3, smaller than that (0.5) of the single factor diffusions. The estimated drift parameters are significant under the GARCH model, indicating mean-reversion. This differs from the estimation results for the U.S. interest rates, where mean reversion becomes weaker after the GARCH effect is introduced (Durham 2003). The specification of drift and diffusion functions affects the estimation of dummy coefficients. The estimated elasticity parameter ρ_D is significantly negative for both the no drift GARCH-CEV and linear drift GARCH-CEV models. However, it becomes insignificant if a nonlinear drift is introduced. Among all GARCH models, the model with nonlinear drift and level effect has the largest robust log-likelihood. The additional contribution of a nonlinear drift is important.

Table 6 reports the estimation results of Markov regime switching models, which show that the spot rate behaves quite differently between regimes. Both regimes show mean reversion for the models with linear drift, with higher and lower long run means respectively. For the linear drift CEV model, the higher long run mean is 5.25% and the lower long run mean is 2.75%. For the linear drift GARCH model, the higher long run mean is 6.78% and the lower long run mean is 2.24%. The model with both CEV and GARCH together has a higher long run mean 5.23% and a lower long run mean 2.73%. All estimated GARCH parameters are significant, and the sum of parameters, $\beta_1 + \beta_2$, is smaller than 1 when the level effect is not included. The level effects in two regimes are significant, with the estimated elasticity parameter about 0.5 in one regime and about 1.5 in the other regime. The level effect elasticity parameters are higher and more stable than the estimation results of the U.S. interest rates. The volatility of one regime in CEV models is about 15, 5 and 5 times of the other for no linear, linear drift and nonlinear drift respectively. For GARCH models, the relative volatility ratio between two regimes is about 5. For CEV-GARCH models, the ratio depends on drift specification. It is about 2 for no drift and 5 for linear and nonlinear drifts. Higher volatility is related to a higher level effect for most specifications except for the no-drift CEV-GARCH model, i.e. the regime with higher volatility has higher dependence on the interest rate level. Compared with GARCH models, the Markov regime switching models have much higher robust log-likelihood, implying the improvement of goodness of fit over GARCH models. The models with level effect performed better than those with GARCH effect. This is also in contrast with the estimation results for the U.S. spot rates (Hong et al. 2004). Interestingly, combining both level effect and GARCH effect together does not improve much the goodness of fit. The models with nonlinear drift have the largest log-likelihood, although some parameters are insignificant.

Table 7 reports the estimation results of discretized jump diffusion models. The mean reversion is still significant, with a long run mean about 2.30%. All GARCH parameter estimates are significant, with the sum $\beta_1 + \beta_2$ smaller than 1. The GARCH parameter estimates are smaller than those of pure GARCH models. Without GARCH effects, the level effect elasticity parameter estimate is more than 1.5. However, with the GARCH effects, the level effect becomes weaker. Apparently, GARCH specifications help capture volatility clustering of the Chinese spot rates. The parameter estimates of jump probability are overwhelmingly significant under GARCH and CEV-GARCH specifications. The specifications of both conditional mean and variance affect the estimation of jump size parameters. The jump size is about 0.65% for the CEV specification, and becomes smaller for GARCH and CEV-GARCH specifications. The volatility parameter estimates in all specifications remain stable at about 1.7%. The drift dummy coefficient estimate α_D is significant in all models, suggesting a higher interest rate level before 1999. The elasticity parameter ρ_D is significant for CEV and becomes insignificant for most CEV-GARCH models, which shows the effectiveness of GARCH effect in capturing volatility clustering of the Chinese spot rates. Both the dummy coefficients for jump probability are significantly negative and reflect a higher jump probability before 1999. Similar to Markov regime switching models, the jump diffusion models with CEV perform a bit better than those with GARCH effects, which is in contrast with the empirical results for the U.S. spot rates. On the other hand, combining both CEV and GARCH effects does not improve much the goodness of fit. Again, the jump diffusion models with a nonlinear drift have the largest log-likelihood, although some parameters are insignificant.

To sum up, our estimation reveals some important stylized facts of the Chinese spot rates:

(1) There exists significant mean reversion in the Chinese spot rates. Although some parameters (but not all) are insignificant, a nonlinear drift specification outperforms a linear drift specification. Ait-Sahalia’s (1996) type nonlinear drift is useful in modeling the Chinese spot rate dynamics. This differs from the empirical evidence for the U.S. spot rates. Furthermore, the specification of conditional mean affects the estimation of other parameters involving GARCH and level effects.

(2) There exists significant conditional heteroscedasticity in the Chinese spot rates, which can be captured by GARCH effect or level effect. Combining both GARCH effect and level effect, however, does not improve much the goodness of fit. The models with level effect generally outperform the models with GARCH effect in terms of robust log-likelihood

value.

(3) Regime switching and jump help capture volatility clustering and especially the excess kurtosis and heavy-tails of the Chinese interest rates, which display frequent extreme changes.

(4) The Chinese spot rates behave significantly differently before and after 1999, when a structure break occurred. The level/volatility of interest rates and the probability of jump probability are significantly higher before 1999. However, the level effect, namely the dependence of the interest rate volatility on the interest rate level becomes stronger after 1999.

(5) There are both significant similarities and differences between the time series dynamics of the Chinese spot rates and the U.S spot rates. As summarized in Table 8, there exist significant mean reversion and conditional heteroskedasticity in both the Chinese and the U.S. spot rates. Regime switching and jump help capture volatility clustering and especially the excess kurtosis and heavy-tails of both the Chinese and the U.S. interest rates. On the other hand, there are also significant differences between the dynamics of the Chinese spot rates and the U.S. spot rates. For single factor diffusion models, the contribution of nonlinear drift beyond a linear drift is significant for the Chinese spot rates, while this is inconclusive for the U.S. spot rates. The elasticity parameter estimate is 0.5 for the Chinese spot rates, and is 1.5 in CKLS (1992) for the U.S. spot rates. For GARCH models, mean reversion is significant when GARCH effect is included for the Chinese spot rates, but it is insignificant for the U.S. spot rates. For Markov regime switching models, there exist significant differences in the estimation results of elasticity, mean reversion, volatility ratios, and the relative performance of level effect and GARCH effect. For jump diffusion models, there also exist significant differences in the estimation results of elasticity, jump size, and the relative performance of CEV effect and GARCH effect.

5.2 Impact of non-robust estimation

Because there are relatively frequent jumps in the Chinese spot rates, we have used a robust MLE method to estimate spot rate models. To examine the impact of non-robust estimation on empirical results, Table 9 reports the estimation results of several spot rate models using the conventional MLE and robust MLE respectively. We choose the nonlinear drift diffusion, nonlinear drift GARCH, nonlinear drift GARCH-regime switching (nonlinear drift GARCH-RS) and nonlinear drift GARCH-jump diffusion (nonlinear drift GARCH-JD) models for illustration.[6] The estimated parameters of nonlinear drift diffusion and nonlinear drift GARCH models using the normal likelihood are different from those using the robust MLE, indicating that the estimation results of single factor diffusion models and GARCH models are significantly affected by outliers, which highlights the importance of robust estimation for these models.

[6] The results of other models are quite similar and not reported here. They are available from the authors upon request.

However, for more sophisticated models such as regime switching models and jump diffusion models, the results of the conventional MLE become quite similar to those of the robust MLE. The difference of their log-likelihood values is also small. This is conceivable since while the outliers in original data are not captured by such simple models as one factor diffusion models, regime switching models and jump diffusion models can effectively capture the impacts of outliers, which make the estimation of other model parameters relatively robust.

6. Model validation

6.1 Portmanteau specification testing

To validate the estimated Chinese spot rate models, we apply the robust Hellinger-metric test described in Section 4 to each class of spot rate models.

Table 10(a) reports the $\hat{W}(p)$ statistics with lag order $p=1,5,10$ respectively for single factor diffusion models. The $\hat{W}(p)$ statistics range from 98.83 to 518.95, suggesting that all eight diffusion models are firmly rejected at any reasonable significance level. The lognormal model performs the worst among the eight models. The $\hat{W}(p)$ values for the CIR model are smaller than those for the Vasicek model, and even smaller for the nonlinear drift model. This suggests that nonlinear drift and level effect help capture the dynamics of the Chinese spot rates. The nonlinear drift CEV model performs the best. Nevertheless, the large $\hat{W}(p)$ statistics for the eight single factor diffusion models indicate that none of them can adequately capture the Chinese interest rate dynamics.

Table 10(b) reports the $\hat{W}(p)$ statistics for GARCH models, which range from 58.52 to 188.23, significantly smaller than those of single factor diffusion models. This highlights the effectiveness of GARCH specification in modeling the Chinese spot rates. Introducing the level effect improves the goodness of fit. The no drift model performs the worst among the six models. Introducing the linear drift substantially reduces the $\hat{W}(p)$ value. The $\hat{W}(p)$ values of nonlinear drift models are smaller than those of linear drift models, implying that they perform better than the linear drift models. The contribution of nonlinear drift beyond the linear drift specification is evident, especially for CEV models. However, all six GARCH models are strongly rejected at any reasonable significance level.

Table 10(c) reports the $\hat{W}(p)$ statistics for Markov chain regime switching models, which range from 12.53 to 45.96, much smaller than those of GARCH models. This indicates the greater flexibility of regime switching models over GARCH models in capturing the Chinese interest rate dynamics. However, they are still overwhelmingly rejected at any reasonable significance level.

The no drift model performs the worst among the nine models. Introducing the linear drift reduces the $\hat{W}(p)$ value substantially, and the nonlinear drift specification gives further improvement. The combination of level effect and GARCH effect also improves the performance of models a little. The models with level effect and with either linear drift or nonlinear drift have smaller $\hat{W}(p)$ values than the models with GARCH effect, suggesting that the level effect better captures the volatility clustering than GARCH models. This is consistent with the relative likelihood values.

Table 10(d) reports the $\hat{W}(p)$ statistics for jump diffusion models, which range from 16.56 to 41.06, which is very similar to Markov regime switching models. Regime switching models and jump diffusion models are two best classes of models to characterize the dynamics of the Chinese spot rates, particularly in capturing the extreme observations. However, the $\hat{W}(p)$ statistics are all overwhelmingly significant at any reasonable significance level, and thus all the nine models are firmly rejected. The no drift model performs the worst among the nine models. Introducing the linear drift substantially reduces the $\hat{W}(p)$ value. There is a marginal improvement when a nonlinear drift is introduced. Combining level effect and GARCH effect does not improve the goodness of fit relative to separate inclusion of CEV effects or GARCH effects.

Due to the existence of frequent extreme observations in Chinese spot rate data, we have used a robust MLE method to estimate spot rate models and a robust Helliger metric-based specification test to validate these models. To examine the impact of non-robust MLE and specification tests on model validation, Table 11 reports the $\hat{W}(p)$ statistics of several spot rate models using the conventional MLE and the quadratic form test.[7] It also reports the $\hat{W}(p)$ statistics using the robust MLE and Hellinger-metric based test. The $\hat{W}(p)$ statistics of the nonlinear drift diffusion model and nonlinear drift GARCH models with the conventional MLE estimation are different from those using the robust MLE. The results of specification tests of single factor diffusion models and GARCH models are significantly affected by the non-robust estimation method. However, for more sophisticated models such as regime switching and jump diffusion models, the $\hat{W}(p)$ statistics using the robust MLE become quite similar to those using the conventional MLE. These findings are consistent with the results of estimation comparisons

[7] The results of other models are quite similar and not reported here. They are available from the authors upon request.

between robust and non-robust estimation. On the other hand, the difference between the $\hat{W}(p)$ statistics based on $\hat{H}(j)$ and $\hat{Q}(j)$ using the same estimation method is substantial for the nonlinear drift diffusion model, becomes much smaller for GARCH, and is quite small for regime switching and jump models. Therefore, while the results of specification test of such spot rate models as single factor diffusion and GARCH models are affected by the non-robust estimation method and specification test, more sophisticated models such as regime switching and jump diffusion models that have more flexibilities in capturing extreme observations have relatively robust results on model validation. Most importantly, the relatively ranking of models revealed by the quadratic form test is the same as that by the Hellinger metric test.

To sum up, our specification tests reveal some important findings in modeling the Chinese spot rates, most of which are consistent with those in estimation. In particular,

(1) A linear drift is significant, and the additional contribution of a nonlinear drift beyond a linear drift is not negligible. For single factor diffusion models, the contribution of a nonlinear drift over a linear drift is significant. When GARCH, regime switching or jump effects are introduced, the contribution of a nonlinear drift over a linear drift becomes less but still significant.

(2) The level effect or GARCH effect can capture volatility clustering of interest rates, and level effect can better capture volatility clustering than GARCH effect. However, there is little improvement in combining both effects together.

(3) Introducing the GARCH/level effect, regime switching effect and jump effect can improve the performance of various models for the Chinese spot rates. However, they are all rejected by the $\hat{W}(p)$ tests at any reasonable significance level, suggesting that they are still grossly misspecified.

6.2 Separate inference

The results of the portmanteau tests suggest that introducing GARCH, regime switching and jump effectively reduce specification errors of Chinese spot rate models but all of them are still strongly rejected. Therefore it may be interesting to examine possible source of model misspecification. For this purpose, we first check the model marginal distribution and then check model dynamics.

The model marginal distribution is characterized by the U[0,1] property of generalized residuals. If the model could characterize the marginal distribution adequately, the histogram of generalized residuals will be more or less horizontal. The closer to a horizontal line, the more adequate marginal distribution specification the model has. Figure 3 plots the histograms of generalized residuals for different models. Panel (a) plots the histograms of generalized residuals for single factor diffusion models, which are far from being uniform, with a high peak around 0.5.

This implies that these diffusion models are inadequate in capturing excess kurtosis. The histograms of generalized residuals for Vasicek, CIR, CKLS and nonlinear drift diffusion models are more uniformed than those for Random walk, lognormal, Dothan and pure CEV models. Panel (b) plots the histograms of generalized residuals for GARCH models. The peak is much lower than that of single factor diffusion models. This reflects the improvement of marginal distribution specification by introducing GARCH effects, which can capture some extreme changes. The models with linear drift and nonlinear drift have lower peak around zero and are closer to uniformity than the models with no drift. Therefore, mean reverting helps improve the fitting of the marginal distribution of Chinese spot rates. However, the histograms of generalized residuals for GARCH models are still different from the uniform distribution.

Panels (c) and (d) plot the histograms of generalized residuals for regime switching models and jump diffusion models respectively. These histograms are very similar to the uniform distribution, with a lower peak around zero for linear and nonlinear drift specifications. These results suggest that regime switching and jump diffusion could effectively model the marginal distribution of the Chinese spot rates, particularly the heavy tails. In summary, drift, GARCH, regime switching and jump all help fit the marginal distribution of the Chinese spot rates.

Next, we examine the dynamics of different models by checking the i.i.d. property of their generalized residuals. If the generalized residuals are i.i.d., then $\rho_{ml}(j)$ will be zero for any choice of *(m,l)* and M*(m,l)* in (4.6) will converge to a N(0,1) distribution. We can compare the M*(m,l)* values of different models to examine the source of dynamic misspecification.

Table 12 reports the M*(m,l)* values of spot rate models for *(m,l)* = (1,1), (1,2), (2,1), (2,2), (3,3) and (4,4). Approximately, M(1,1) checks autocorrelations in level, M(1,2) checks the ARCH-in-mean effect, M(2,1) checks the leverage effect, and M (2,2), M(3,3) and M(4,4) check autocorrelations in higher order moments. Several findings are recorded. First, mean reverting characterized by linear drift and nonlinear drift help modeling the dynamics by reducing the values of M*(m,l)*. Most of M*(m,l)* values become smaller after linear and nonlinear drift are introduced. Interestingly, more sophisticated models, such as regime switching and jumps do not help reduce the autocorrelations in level as measured by M(1,1). Second, the M(1,2) values of single factor models vary a lot, ranging from 0.78 (Lognormal model) to 18.12 (RW model). The M(1,2) values of most GARCH models are smaller than those of single factor diffusion models, while the M(1,2) values of regime switching and jump models are similar to those of GARCH models. The M(1,2) values of some GARCH models are even insignificant. Third, the M(2,1) values of single factor models also vary a lot, ranging from 3.62 (Dothan model) to 104.21 (Vasicek model). The M(2,1) values of GARCH models are generally smaller than those of single factor diffusion models, and becomes even smaller when regime switching and jumps are introduced. The M(2,1) values of some regime switching and jump models are insignificant. Therefore, GARCH, regime switching and jump all help modeling the asymmetric features in volatility. Fourth, the M(2,2) and M(4,4) values of GARCH models are smaller than those of single factor diffusion models. The regime switching and jump models with level effect have similar M(2,2) and M(4,4) values to GARCH models, while the regime switching and jump models with GARCH effect have much smaller values of M(2,2) and M(4,4). This means that the GARCH specification captures the autocorrelation in second moments and forth moments more effectively than the CEV specification in sophisticated models. Finally, the M(3,3) values of GARCH models are smaller than those of single factor diffusion models, while regime switching and jump make no improvement in reducing the values of M(3,3).

To sum up, our separate inference reveals some important findings in modeling the marginal distribution and dynamics of the Chinese spot rates. Linear drift, nonlinear drift and GARCH models reduce specification errors in both the marginal distribution and dynamics of the Chinese spot rates. Regime switching and jump models also reduce specification errors in both dimensions, but their improvement in marginal distribution is more significant than in model dynamics.

7. Conclusion

Based on a daily sample of the 7-day Chinese spot interest rates, we estimate and test a variety of spot rate models, which include discretized single factor diffusion models, GARCH models, Markov regime switching models and jump diffusion models. To alleviate possible impact of frequent outliers in the Chinese spot rates, a robust M-estimation method and a robust Hellinger metric-based test are used.

We document that introducing GARCH effects significantly improves the goodness of fit. Regime switching and jump effects help capturing volatility clustering and especially the excess kurtosis and heavy tails of the Chinese spot interest rates. Moreover, there exists significant mean reverting, and the contribution of nonlinear drift is significant. Level effect is also significant.

Although GARCH, regime switching and jumps are important for modeling the Chinese spot rate dynamics, they are still grossly misspecified. There is a long way to go before we reach a correct specification for the Chinese spot rate dynamics. We further explore possible sources of model misspecification by examining the marginal distribution and model dynamics separately. We find that linear drift, nonlinear drift and GARCH models reduce specification errors in both marginal distribution and dynamics. Regime switching and jump models also reduce specification errors in both dimensions, but their improvement in marginal distribution is more significant than in model dynamics, apparently due to their ability to capture extreme observations. This may have useful implications for further modeling the Chinese spot rates.

References:

Ahn, D.H., Gao, B., 1999. A parametric nonlinear model of term structure dynamics. Review of Financial Studies 12, 721-762.

Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385-426.

Ait-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361-1395.

Andersen, T.G., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 77, 343-377.

Ang, A., Bekaert, G., 2002. Regime switches in interest rates. Journal of Business and Economic Statistics 20, 163-182.

Balduzzi, P., Eom, Y., 2000. Non-linearities in US Treasury rates: a semi-nonparametric approach. Working paper, Boston College.

Ball, C.A., Torous, W.N., 1998. Regime shifts in short term riskless interest rates. Working paper, Anderson Graduate School of Management at UCLA.

Bandi, F.M., 2002. Short-term interest rate dynamics: a spatial approach. Journal of Financial Economics 65, 73-110.

Bansal, R., Zhou, H., 2002. Term structure of interest rates with regime shifts. Journal of Finance 57, 1997-2043.

Baz, J., Das, S.R., 1996. Analytical approximation of the term structure for jump-diffusion process: a numerical analysis. Journal of Fixed Income 6, 78-86.

Brenner, R.J., Harjes, R.H., Kroner, K.F., 1996. Another look at models of the short-term interest rate. Journal of Financial and Quantitative Analysis 31, 85-107.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209-1227.

Chapman, D.A., Pearson, N.D., 2000. Is the short rate drift actually nonlinear? Journal of Finance 55, 355-388.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525-577.

Czellar, V., Karolyi, G.A., Ronchetti, E., 2007. Indirect robust estimation of the short-term interest rate process. Journal of Empirical Finance 14, 546-563.

Dai, Q., Singleton, K.J., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943-1978.

Das, S.R., 2002. The surprise element: jumps in interest rates. Journal of Econometrics 106, 27-65.

Dell'Aquila, R., Ronchetti, E., Trojani, F., 2003. Robust GMM analysis of models for the short rate process. Journal of Empirical Finance 10, 373-397.

Durham, G.B., 2003. Likelihood-based specification analysis of continuous-time models of the short-term interest rate. Journal of Financial Economics 70, 463-487.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business and Economic Statistics 20, 297-338.

Elerian, O., Chib, S., Shephard, N., 2001. Likelihood Inference for discretely observed nonlinear diffusions. Econometrica 69, 959-993.

Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357-384.

Hong, Y., 1996, Consistent testing for serial correlation of unknown form. Econometrica 64, 837-864.

Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies 18, 37-84.

Hong, Y., Li, H., Zhao, F., 2004. Out-of-sample performance of discrete-time spot interest rate models. Journal of Business and Economic Statistics 22, 457-474.

Hong, Y., Li, H., Zhao, F., 2007. Can the random walk model be beaten in out-of-sample density forecasts? evidence from intraday foreign exchange rates. Journal of Econometrics 141, 736-776.

Huber, P.J., 1981. Robust Statistics. Wiley-Interscience.

Johannes, M., 2004. The statistical and economic role of jumps in continuous-time interest rate models. Journal of Finance 59, 227-260.

Jones, C.S., 2003. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies 16, 793-843.

Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347-370.

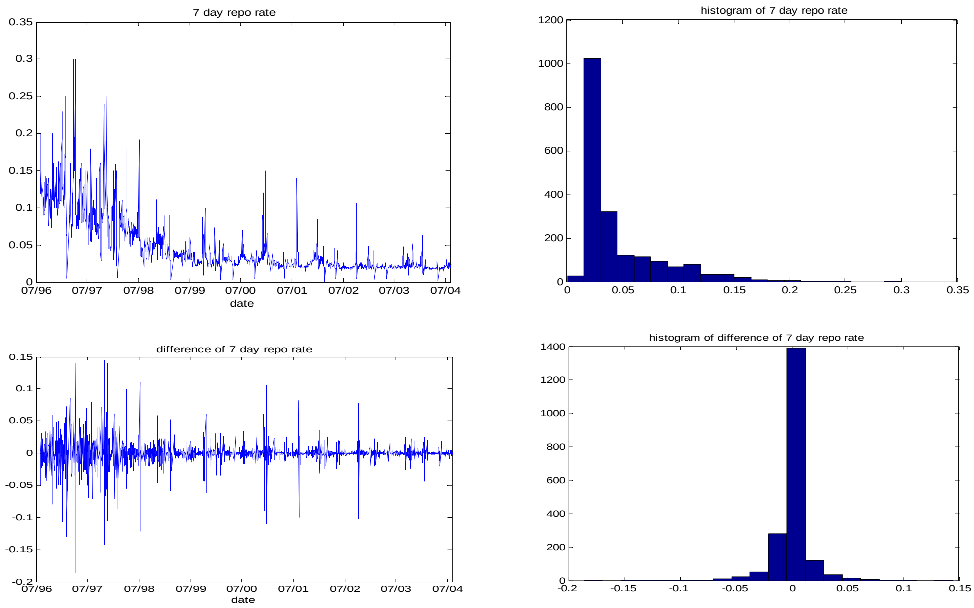
Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Review of Financial Studies 11, 449-487.

Sanders, A.B., Unal, H., 1988. On the intertemporal behavior of the short-term rate of interest. Journal of Financial and Quantitative Analysis 23, 417-423.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley-Interscience.

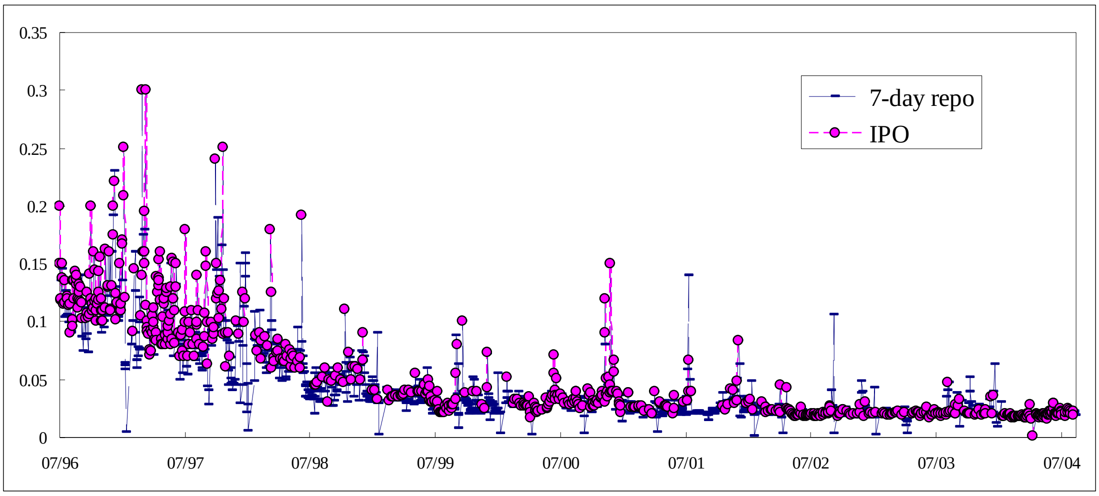
Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973-2002.

Figure 1. Daily 7-day repo rates between July 22, 1996 and August 26, 2004. This figure plots the level and change series of daily data as well as their histograms.



Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) The first row plots the level series with the second row plots the change series.

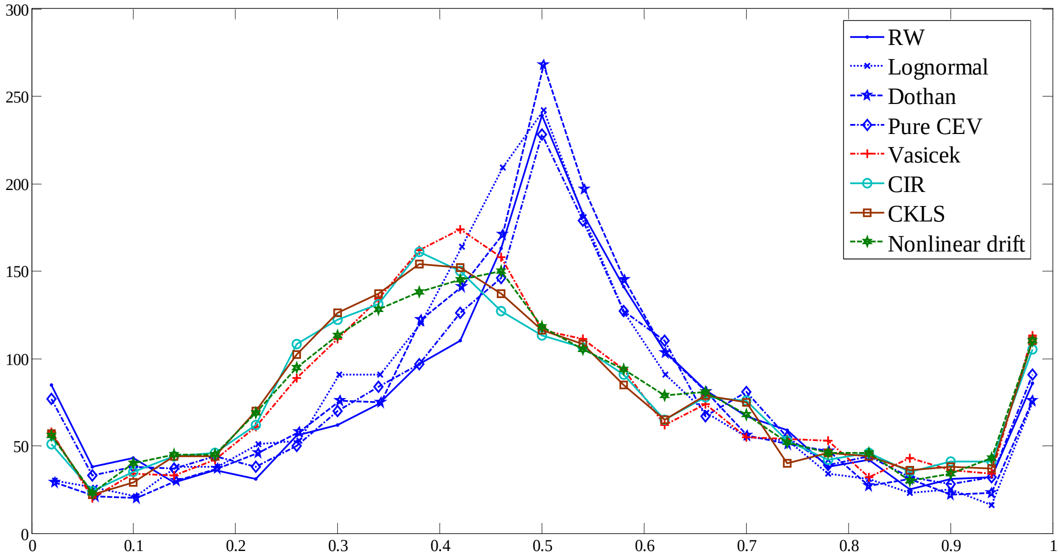
Figure 2. The dynamics of 7-day repo and IPO. This figure plots the dynamics of 7-day repo rates with IPO during the sample period.



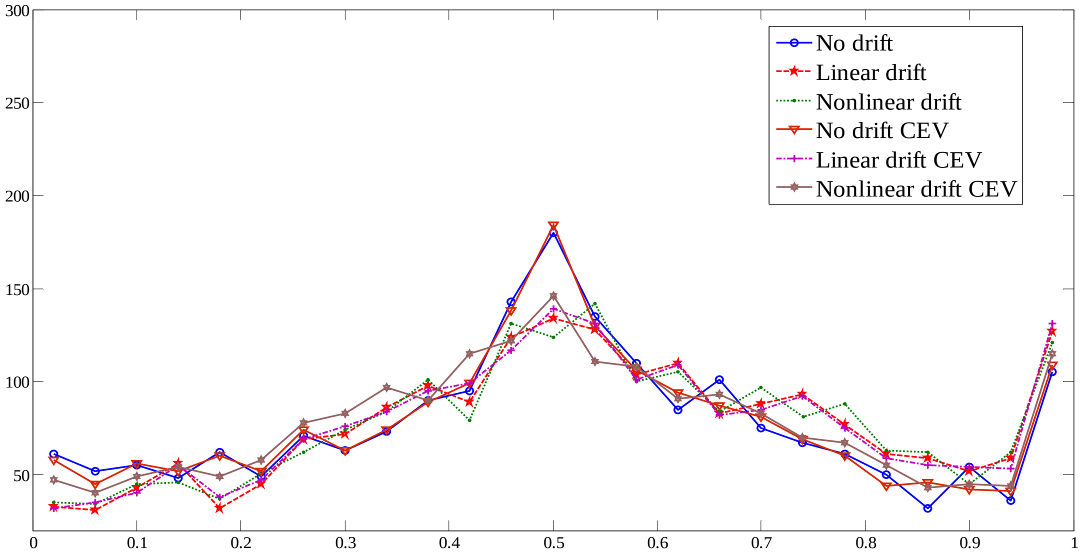
Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) the round points are those days with IPO.

Figure 3. Histograms of generalized residuals. This figure plots the histograms of generalized residuals of discrete time spot rate models.

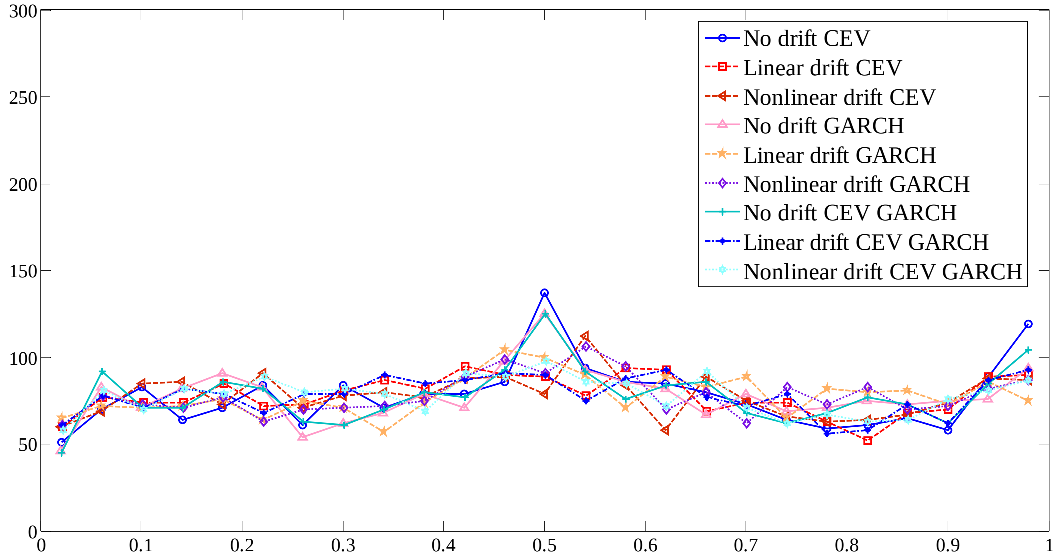
(a) Single factor diffusion models



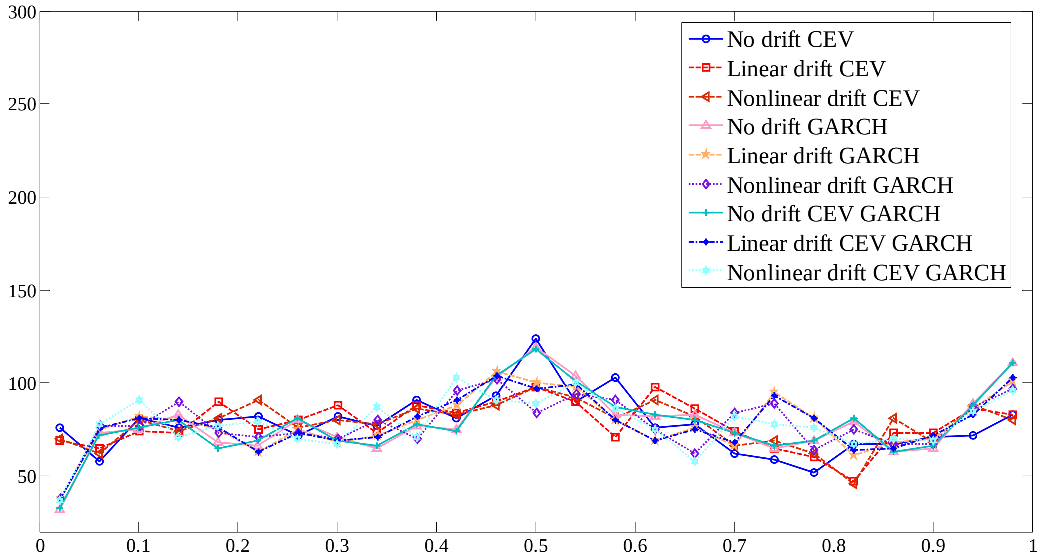
(b) GARCH models



(c) Regime switching models



(d) Jump diffusion models



Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) panel (a), (b), (c), (d) plot the histograms of generalized residuals for single factor diffusion, GARCH, Markov regime switching and jump diffusion models respectively.

Table 1. Characteristics of Chinese interest rate liberalization. This table summarizes the history and recent reforms of Chinese interest rate liberalization.

<table>
<tr><td rowspan="4">History:
● Market segmentation;
● Strict regulation</td><td rowspan="2">Short term</td><td>Inter-bank borrowing</td></tr>
<tr><td>Repurchase: Two segmented sub-markets
● Inter-bank repurchase market
● Exchange repurchase market: serious impact by IPO.</td></tr>
<tr><td>Middle term</td><td>Saving rates strictly regulated by Chinese central bank</td></tr>
<tr><td>Long term</td><td>Two segmented sub-markets:
● Inter-bank long term bond market
● Exchange long term bond market</td></tr>
<tr><td>Reforms:
● In a stable process</td><td colspan="2">● Issue bonds at both inter-bank market and exchange market
● Permit some eligible securities and trust companies to join the issuing
● Propose the market maker system in 2001 on secondary market
● Issue bonds systematically ranging from long term to short term
● Propose many other instruments, such as Stock Index Futures, Bond Futures and Monetary Market Fund (MMF).</td></tr>
</table>

Note: This table outlines the history and recent reforms of Chinese interest rate liberalization.

Table 2. Summary of Chinese short tem money trading. This table reports the yearly trading volume of Chinese short term money market between 1997 and 2004.

	Repo				Inter-bank borrowing
Year	1-day repo (Billion RMB)	7-day repo (Billion RMB)	14-day repo (Billion RMB)	1-month repo (Billion RMB)	All (Billion RMB)
1997	0	75.41	20.86	12.87	
1998	0	89.40	25.37	24.94	98.95
1999	0	168.29	13.72	22.44	329.16
2000	0	619.64	146.17	129.37	672.81
2001	0.68	1654.42	303.91	158.33	808.20
2002	3.19	3650.32	609.22	240.18	1210.72
2003	256.41	9291.20	1347.72	370.97	2411.34
2004	3615.31	9257.76	1764.23	582.71	1455.55

Notes: (1) The trading volume of repo market is the total trading volume in both exchange markets and inter-bank markets; (2) The trading volume for inter-bank borrowing market is the total trading volume of inter-bank borrowing with all maturities, including 1 day, 7 days, 14 days, 1 month and 4 months.

Table 3. Spot rate models considered for evaluation. This table lists the spot rate models that will be evaluated in paper.

Model	Mean	Volatility
(a) Discretized single factor diffusion models		
Random walk	α_0	σ
Lognormal	$\alpha_1 r_{t-1}$	σr_{t-1}
Dothan	0	σr_{t-1}
Pure CEV	0	σr_{t-1}^{ρ}
Vasicek	$\alpha_0 + \alpha_1 r_{t-1}$	σ
CIR	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{0.5}$
CKLS	$\alpha_0 + \alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	σr_{t-1}^{ρ}
(b) GARCH models		
No drift GARCH	0	$\sigma \sqrt{h_t}$
Linear drift GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma \sqrt{h_t}$
Nonlinear drift GARCH,	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma \sqrt{h_t}$
No drift CEV-GARCH	0	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$
Linear drift CEV-GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$
Nonlinear drift CEV GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$
(c) Markov regime switching models		
No drift RS CEV	0	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Linear drift RS CEV	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Nonlinear drift RS CEV,	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
No drift RS GARCH	0	$\sigma(s_t)\sqrt{h_t}$
Linear drift RS GARCH	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t)\sqrt{h_t}$
Nonlinear drift RS GARCH	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t)\sqrt{h_t}$
No drift RS CEV GARCH	0	$\sigma(s_t) r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
Linear drift RS CEV GARCH	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
Nonlinear drift RS CEV GARCH	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
(d) Discretized jump diffusion models		
No drift JD CEV	0	σr_{t-1}^{ρ}
Linear drift JD CEV	$\alpha_0 + \alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift JD CEV,	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	σr_{t-1}^{ρ}
No drift JD GARCH	0	$\sigma \sqrt{h_t}$
Linear drift JD GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma \sqrt{h_t}$
Nonlinear drift JD GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma \sqrt{h_t}$
No drift JD CEV GARCH	0	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$
Linear drift JD CEV GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$
Nonlinear drift JD CEV GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma r_{t-1}^{\rho}\sqrt{h_t}$

Notes: panel (a), (b), (c), (d) list 8 specifications of discretized single factor diffusion models, 6 specifications of GARCH models, 9 specifications of Markov regime switching models, and 9 specifications of jump diffusion models that will be evaluated respectively.

Table 4. Parameter estimates for the single factor diffusion models. This table reports the robust parameter estimates of single factor diffusion models with respective robust standard errors and robust log-likelihood value.

Parameters	RW	Lognormal	Dothan	PCEV	Vasicek	CIR	CKLS	Nonlinear drift
α_{-1}								2.7E-5
								(3.00E-06)
α_{0}	-5.1E-05	0.0119			7.31E-03	7.63E-03	7.59E-03	3.79E-03
	(1.79E-04)	(0.0089)			(4.55E-04)	(4.50E-04)	(4.59E-04)	(8.30E-04)
α_{1}					-0.2858	-0.2969	-0.2957	-0.1813
					(0.0164)	(0.0175)	(0.0177)	(0.0362)
α_{2}								-0.4538
								(0.2432)
σ	0.0063	0.3214	0.3213	0.0316	0.0059	0.0358	0.0323	0.0458
	(0.0002)	(0.0079)	(0.0079)	(0.0023)	(0.0001)	(0.0008)	(0.0026)	(0.0042)
ρ				0.4384			0.4660	0.5792
				(0.0189)			(0.0211)	(0.0243)
α_D	-1E-06	0.0079			0.0171	0.0160	0.0170	0.0143
	(9.76E-04)	(0.0155)			(1.39E-03)	(1.22E-03)	(1.35E-03)	(1.27E-03)
σ_D	0.0168	-0.0188	-0.0187		0.0166	0.0384		
	(0.0008)	(0.0130)	(0.130)		(0.0008)	(0.0026)		
ρ_D				-0.2733			-0.2885	-0.2571
				(0.0185)			(0.0189)	(0.0188)
Log-likelihood	5989.75	5454.11	5452.02	6066.62	6136.56	6222.39	6210.99	6320.99

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimation; (3) The reported values are the estimated parameters with their standard errors in brackets; (4) The models are nested by: $\Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t, \{z_t\} \sim iid.N(0,1)$.

Table 5. Parameter estimates for GARCH models. This table reports the robust parameter estimates of GARCH models with respective robust standard errors and robust log-likelihood value.

Parameters	No drift	linear drift	Nonlinear drift	No drift CEV	linear drift CEV	Nonlinear drift CEV
α_{-1}			4.20E-05			4.1E-05
			(7.0E-06)			(6.00E-06)
α_{-0}		3.03E-03	-1.21E-03		3.03E-03	-1.34E-03
		(3.43E-04)	(8.34E-04)		(3.44E-04)	(7.15E-04)
α_{1}		-0.1371	-0.0194		-0.1351	2.13E-03
		(0.054)	(0.0324)		(0.0153)	(0.0284)
α_{-2}			-0.6638			-1.0803
			(0.2608)			(0.2333)
ρ				0.1732	0.1043	0.2501
				(0.0308)	(0.0362)	(0.0373)
β_0	1.60E-06	1.78E-06	1.58E-06	6.51E-06	4.35E-06	9.27E-06
	(2.91E-07)	(2.96E-07)	(2.79E-07)	(2.09E-06)	(1.54E-06)	(3.50E-06)
β_1	0.3909	0.5029	0.4692	1.3058	1.0186	1.8787
	(0.0367)	(0.0534)	(0.0515)	(0.3002)	(0.2644)	(0.5171)
β_2	0.5793	0.5063	0.5333	0.5770	0.5001	0.6269
	(0.0251)	(0.0285)	(0.0291)	(0.0267)	(0.0292)	(0.0312)
α_D		2.16E-03	1.22E-03		2.20E-3	7.31E-03
		(7.36E-04)	(8.06E-04)		(7.50E-04)	(1.01E-03)
σ_D	0.3529	0.2907	0.2830			
	(0.0692)	(0.0656)	(0.0659)			
ρ_D				-0.05	-0.0725	-0.0162
				(0.0207)	(0.0208)	(0.0243)
Log-likelihood	6542.02	6614.06	6636.98	6551.98	6618.87	6652.66

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimation; (3) The reported values are the estimated parameters with their standard errors in brackets; (4) The GARCH models are nested by: $\Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t, , h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-2}^{2\rho} z_{t-1}^2)$, $\{z_t\} \sim iid.N(0,1)$.

Table 6. Parameter estimates for Markov regime switching models. This table reports the robust parameter estimates of Markov regime switching models with respective robust standard errors and robust log-likelihood value.

Parameters	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
α_{-1} (1)			8.10E-05			3.30E-05			1.90E-05
			(1.00E-06)			(2.30E-05)			(6.00E-06)
α_{0} (1)		1.27E-3	-5.10E-03		6.01E-03	-4.86E-03		0.018	9.37E-03
		(2.56E-04)	(4.59E-04)		(1.22E-03)	(2.61E-03)		(1.52E-03)	(4.24E-03)
α_{1} (1)		-0.0461	0.0872		-0.0886	0.3299		-0.3444	-0.1261
		(0.0111)	(0.0233)		(0.0211)	(0.0801)		(0.0287)	(0.1607)
α_{-2} (1)			-0.8287			-2.7001			-0.860
			(0.2374)			(0.4909)			(0.740)
α_{-1} (2)			3.00E-06			6.50E-05			1.57E-04
			(4.00E-06)			(9.00E-06)			(3.00E-06)
α_{0} (2)		0.0192	0.0267		1.31E-03	-2.49E-03		1.29E-03	-0.0114
		(1.81E-03)	(3.05E-03)		(3.00E-04)	(9.20E-04)		(2.50E-04)	(7.7E-04)
α_{1} (2)		-0.3658	-0.9181		-0.0586	-0.0415		-0.0473	0.2192
		(0.0669)	(0.1786)		(0.0135)	(0.0308)		(0.0106)	(0.0368)
α_{-2} (2)			8.7605			0.6684			-1.180
			(2.3690)			(0.2392)			(0.369)
ρ (1)	1.4936	1.4629	1.4927	0	0	0	0.1913	0.4526	0.4403
	(0.0443)	(0.0454)	(0.0386)				(0.0466)	(0.0509)	(0.0678)
ρ (2)	0.2107	0.4656	0.3518	0	0	0	0.5461	1.4305	1.4013
	(0.0413)	(0.0524)	(0.0682)				(0.2514)	(0.0625)	(0.0701)
σ(1)	0.6351	0.5596	0.5471	1	1	1	1	1	1
	(0.0970)	(0.0884)	(0.0601)						
			To be continued			**To be continued**			

$\sigma(2)$	0.0474	0.1046	0.1263	0.2180	0.2334	0.2247	0.6753	5.1442	5.1433
	(7.85E-03)	(0.0201)	(0.0291)	(0.0123)	(0.0144)	(0.0128)	(0.4625)	(1.4831)	(0.7657)
β_0				6.7E-06	6.00E-06	6.10E-06	7.25E-05	5.12E-04	4.77E-04
				(1.5E-06)	(1.30E-06)	(1.40E-06)	(5.52E-05)	(3.24E-04)	(3.08E-04)
β_1				0.0816	0.0938	0.0917	0.5197	0.8908	1.1854
				(0.0146)	(0.0170)	(0.0161)	(0.3604)	(0.8399)	(1.1346)
β_2				0.7806	0.7699	0.7725	0.8400	0.9418	0.9205
				(0.0207)	(0.0228)	(0.0224)	(0.0264)	(0.0302)	(0.0359)
c_1	-3.5803	-2.9320	-2.5976	-0.2948	-0.4043	-0.2504	-0.7144	-0.8960	-0.9414
	(0.3595)	(0.2591)	(0.2051)	(0.2563)	(0.2703)	(0.3342)	(0.3761)	(0.3186)	(0.3228)
d_1	28.66	5.6349	4.2790	-7.7933	-8.9769	-17.1658	-5.1214	24.16	10.39
	(11.26)	(6.1755)	(3.840)	(2.5706)	(2.9188)	(4.4883)	(5.5216)	(55.41)	(58.66)
c_2	-1.4936	-0.8422	-1.3570	-2.9761	-3.1409	-2.9895	-2.9628	-2.9349	-2.8205
	(0.0443)	(0.3060)	(0.3106)	(0.2396)	(0.2625)	(0.3168)	(0.2731)	(0.2354)	(0.2521)
d_2	16.42	0.6974	20.62	27.98	29.4765	26.6882	20.2723	66.80	78.99
	(8.17)	(5.0356)	(6.00)	(4.7073)	(5.6900)	(7.4948)	(8.74)	(48.04)	(63.24)
Log-likelihood	6900.88	6957.64	6990.89	6864.33	6893.08	6952.67	6890.09	6964.51	7008.31

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimation; (3) The reported values are the estimated parameters with their standard errors in brackets; (4) The Markov regime switching models are nested by:

$$\begin{cases} \Delta r_t = \alpha_{-1}(s_t) r_{t-1}^{-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho(s_t)} \sqrt{h_t} z_t, \\ h_t = \beta_0 + \beta_1 E[e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2})]^2 + \beta_2 h_{t-1}, \\ e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2}) = [\Delta r_{t-1} - E(\Delta r_{t-1} \mid r_{t-2}, s_{t-1})] / \sigma(s_{t-1}), \\ \{z_t\} \sim iid.N(0,1). \end{cases}$$

Table 7. Parameter estimates for jump diffusion models. This table reports the robust parameter estimates of jump diffusion models with respective robust standard errors and robust log-likelihood value.

Parameters	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
α_{-1}			4.200E-05			5.50E-05			6.2E-05
			(1.10E-05)			(1.00E-06)			(7.00E-06)
α_{-0}		3.50E-03	2.80E-05		2.53E-03	-2.88E-03		2.57E-03	-3.10E-03
		(3.35E-04)	(1.30E-03)		(2.88E-04)	(9.71E-04)		(2.96E-04)	(7.94E-04)
α_1		-0.1524	-0.0873		-0.1119	0.0347		-0.1138	0.0293
		(0.0155)	(0.0466)		(0.013)	(0.0302)		(0.0134)	(0.0273)
α_{-2}			0.0481			-0.8476			-0.7747
			(0.3739)			(0.2366)			(0.2111)
ρ	2.0846	2.0345	2.0641				-0.0289	0.0312	0.2382
	(0.0695)	(0.0777)	(0.0779)				(0.0413)	(0.0439)	(0.0493)
σ	6.1968	4.8305	5.2631						
	(1.5748)	(1.4131)	(1.5431)						
β_0				2.40E-06	2.30E-06	2.20E-06	1.90E-06	2.92E-06	1.55E-05
				(2.84E-07)	(2.78E-07)	(3.38E-07)	(6.58E-07)	(1.07E-06)	(7.74E-06)
β_1				0.4185	0.4108	0.3476	0.3494	0.5282	1.5533
				(0.0540)	(0.0536)	(0.0509)	(0.1125)	(0.1743)	(0.6340)
β_2				0.0808	0.0932	0.1232	0.0829	0.0932	0.1085
				(0.0309)	(0.0312)	(0.0635)	(0.0325)	(0.0309)	(0.0403)
c	-0.2421	1.0533	1.1217	5.6409	5.1219	5.2776	5.7044	5.0519	4.6651
	(0.6108)	(0.5624)	(0.5568)	(0.4893)	(0.4851)	(0.4776)	(0.4939)	(0.4957)	(0.4555)
		To be continued				**To be continued**			

d	106.60	42.5767	37.7181	-136.57	-115.75	-122.4949	-138.1920	-112.6174	-101.1904
	(30.56)	(26.3429)	(26.0954)	(16.20)	(16.8471)	(16.3914)	(16.3180)	(17.3012)	(16.1559)
ψ	0.0062	0.0065	0.0053	-0.0018	0.0011	0.0010	-0.0020	0.0012	0.0048
	(0.0011)	(0.0010)	(0.0010)	(0.0008)	(0.0011)	(0.0012)	(0.0008)	(0.0011)	(0.0012)
γ	0.0164	0.0160	0.0155	0.0181	0.0177	0.0179	0.0184	0.0177	0.0178
	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
α_D		5.42E-03	4.42E-03		5.08E-03	4.33E-03		5.15E-03	4.36E-03
		(6.28E-04)	(7.70E-04)		(5.70E-04)	(5.86E-04)		(5.74E-04)	(6.37E-04)
ρ_D	0.4915	0.5129	0.4879				-0.0562	-0.0128	-0.0217
	(0.0610)	(0.0712)	(0.0704)				(0.0377)	(0.0357)	(0.0401)
σ_D				0.1672	0.0929	0.1360			
				(0.1134)	(0.1101)	(0.1173)			
c_D	-2.0229	-2.7773	-3.2153	-3.2476	-1.9024	-2.1037	-3.2108	-1.8921	-1.8079
	(0.8038)	(0.7418)	(0.7672)	(0.8502)	(0.9728)	(0.9431)	(0.8711)	(0.9837)	(0.9570)
d_D	-78.2167	-31.7124	-18.3584	98.1772	62.2270	71.1090	98.6509	59.6433	56.1294
	(31.4577)	(27.0591)	(27.1862)	(18.6029)	(21.1110)	(20.3435)	(18.6904)	(21.4126)	(19.5480)
Log-likelihood	6906.52	6953.66	6968.03	6886.03	6938.52	6967.65	6885.96	6938.55	6972.94

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimation; (3) The reported values are the estimated parameters with their standard errors in brackets; (4) The jump diffusion models are nested by:

$$\begin{cases} \Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t + J(\psi, \gamma^2)\pi(q_t), \\ h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1} \mid r_{t-2})]^2 + \beta_2 h_{t-1}, \\ \{z_t\} \sim iid.N(0,1), \\ \{\pi(q_t)\} \sim iid.Bernoulli(q_t), \\ J \sim N(\psi, \gamma^2). \end{cases}$$

Table 8. Similarities and differences of time series dynamics of the Chinese spot rates and the U.S. spot rates. This table reports important similarities and differences between the time series dynamics of the Chinese spot rates and the U.S. spot rates.

Models	Similarities	Differences	
	China/U.S.	China	U.S.
Single factor diffusion models	**1.** There exists significant mean reversion. **2.** There exists significant conditional heteroscedasticity, which can be captured by GARCH effect or level effect. **3.** Regime switching and jump help capture volatility clustering and especially the excess kurtosis and heavy-tails of interest rate data.	**1.** The contribution of nonlinear drift is significant; **2.** The estimate of elasticity is about 0.5.	**1**. the contribution of nonlinear drift is ambiguous; **2**. Elasticity estimate: CKLS (1992)-1.5, Hong, Li and Zhao (2004)-0.25;
GARCH models		Mean reversion is still significant after the introduction of GARCH	Mean reversion decreases rapidly after introduction of GARCH
Markov regime switching models		**1.**the elasticity for two regimes are 1.5 and 0.5; **2.** mean reversion is still significant in two regimes; **3**. the volatility ratios are unstable in two regimes: it is about 5 times for GARCH models and unstable for CEV and CEV-GARCH models; **4**. The relationship between volatility and level effect is relatively stable: higher volatility is related to stronger level effect except for no linear CEV-GARCH model; **5.** CEV models have larger likelihood value than GARCH models	**1**. the elasticity for two regimes are 0.8 and 0.1; **2**. mean reversion is significant in only one regime; **3**. the volatility ratios are relatively stable in two regimes: for CEV models it is about 30 times, for GARCH models it is about 4 times, for CEV-GARCH models it is about 3 times; **4**. The relationship between volatility and level effect is relatively unstable: for CEV models higher volatility is related to weaker level effect, for CEV-GARCH models higher volatility is related to stronger level effect; **5.** GARCH models have larger likelihood value than CEV models.
Jump diffusion models		**1.** The elasticity is 1.5 without GARCH effect and decreases to about 0.2 with GARCH effect; **2.** The jump size for GARCH models is smaller than that of CEV models; **3.** CEV models have larger likelihood value than GARCH models	**1.** The elasticity is 0.9 without GARCH effect and decreases to about 0.1 with GARCH effect; **2.** The jump size for GARCH models is larger than that of CEV models; **3.** GARCH models have larger likelihood value than CEV models.

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimations.

Table 9. Comparisons of estimation results using conventional MLE and robust MLE. This table reports the estimation results of several spot rate models using conventional MLE and robust MLE respectively.

Parameters	Nonlinear drift		Noninear drift GARCH		Noninear drift GARCH-RS		Noninear drift GARCH-JD	
Estimation	N	M	N	M	N	M	N	M
α_{-1} (1)	2.2E-05	2.7E-05	4.70E-05	4.20E-05	2.7E-05	3.30E-05	4.6E-05	5.50E-05
α_{-0} (1)	6.799E-03	3.79E-03	-3.40E-03	-1.21E-03	-5.08E-04	-4.86E-03	-1.72E-03	-2.88E-03
α_{1} (1)	-0.2818	-0.1813	0.0845	-0.0194	0.3911	0.3299	1.59E-04	0.0347
α_{-2} (1)	-0.0936	-0.4538	-1.1289	-0.6638	-3.0240	-2.7001	-0.8377	-0.8476
α_{-1} (2)					6.2E-05	6.50E-05		
α_{-0} (2)					-2.18E-04	-2.49E-03		
α_{1} (2)					-0.0497	-0.0415		
α_{-2} (2)					0.6761	0.6684		
ρ (1)	0.5177	0.5792			0	0		
ρ (2)					0	0		
σ(1)	0.0532	0.0458			1	1		
σ(2)					0.2102	0.2247		
β_0			1.83E-06	1.58E-06	7.11E-06	6.10E-06	1.15E-06	2.20E-06
β_1			0.6296	0.4692	0.1098	0.0917	0.2514	0.3476
β_2			0.6416	0.5333	0.7676	0.7725	0.4913	0.1232
c_1					0.4781	-0.2504	4.9892	5.2776
c_2					-17.2068	-17.1658	-92.7187	-122.4949
d_1					-3.1915	-2.9895		
d_2					30.0579	26.6882		
ψ							0.0028	0.0010
γ							0.0260	0.0179
α_D	0.0183	0.0143	-3.86E-03	1.22E-03			5.18E-03	4.33E-03
ρ_D	-0.2173	-0.2571						

σ_D			0.1039	0.2830			0.0818	0.1360
c_D							-1.5609	-2.1037
d_D							56.1302	71.1090
Log-likelihood	5995.76	6320.99	6306.81	6636.98	6929.35	6952.67	6931.18	6967.65

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) N means the conventional MLE while M means the robust MLE; (3) the reported values are the parameter estimates.

Table 10. $\hat{W}(p)$ stats of discrete spot rte models. This table reports the portmanteau statistic $\hat{W}(p)$ of spot rate models.

(a) Discretized single factor diffusion models

p	RW	Lognormal	Dothan	PCEV	Vasicek	CIR	CKLS	Nonlinear drift
1	152.16	183.34	179.12	134.13	127.48	103.86	110.35	98.83
5	305.04	372.69	368.05	269.85	263.62	224.37	237.73	212.89
10	416.22	518.95	511.22	369.60	363.16	310.81	328.08	294.55

(b) GARCH models

p	No drift	Linear drift	Nonlinear drift	No drift CEV	linear drift CEV	Nonlinear drift CEV
1	77.09	72.14	70.85	74.39	72.06	58.52
5	138.54	134.75	131.00	136.25	133.88	114.11
10	188.23	186.14	181.16	186.53	184.15	159.68

(c) Markov regime switching models

p	No drift CEV	Linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
1	29.94	16.43	13.69	26.86	20.23	16.79	25.28	15.00	12.53
5	37.02	18.12	15.26	30.66	20.22	17.73	28.03	16.01	13.10
10	45.96	21.06	17.63	35.42	22.59	20.63	32.67	18.73	16.55

(d) Jump diffusion models

p	No drift CEV	Linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
1	32.33	16.50	17.46	32.29	22.21	17.40	31.67	22.21	16.56
5	36.01	20.93	20.83	41.87	26.55	20.71	40.76	26.78	18.02
10	41.06	24.95	23.42	51.01	31.99	24.32	49.51	32.41	21.17

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) A robust M-estimation method is used in parameter estimation; (3) The test statistic based on the square Hellinger metric is used in calculating the portmanteau test statistic, which is expected to effectively reduce the impact of outliers; (4) Upper-tailed N(0,1) critical value (e.g., 1.645 at 5% level) is used for specification tests.

Table 11. Comparisons of model validation based on different estimation method and specification tests. This table reports the portmanteau test results of several spot rate models using conventional MLE and the quadratic form test. It also reports the portmanteau test results using robust MLE and the Hellinger metric test.

	Lag	Conventional MLE		Robust MLE	
		$\hat{W}(p)$ based on $\hat{H}(j)$	$\hat{W}(p)$ based on $\hat{Q}(j)$	$\hat{W}(p)$ based on $\hat{H}(j)$	$\hat{W}(p)$ based on $\hat{Q}(j)$
Nonlinear drift	1	161.71	287.81	98.83	140.07
	5	351.76	592.54	212.89	280.44
	10	494.30	795.06	294.55	373.51
Nonlinear drift GARCH	1	122.63	136.54	70.85	61.03
	5	257.32	273.11	131.00	121.45
	10	362.31	369.54	181.16	167.14
Nonlinear drift GARCH-RS	1	19.35	19.64	16.79	17.31
	5	22.87	24.22	17.73	18.45
	10	27.83	28.84	20.63	21.26
Nonlinear drift GARCH-JD	1	21.51	20.35	17.40	14.29
	5	29.18	28.36	20.71	17.60
	10	36.04	35.52	24.32	21.46

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) $\hat{W}(p)$ based on $\hat{H}(j)$ uses squared Hellinger metric, which is expected to effectively reduce the impact of outliers; (3) $\hat{W}(p)$ based on $\hat{Q}(j)$ uses quadratic form, which is not robust to extreme observations; (4) Upper-tailed N(0,1) critical value (e.g., 1.645 at 5% level) is used for specification tests.

Table 12. M*(m,l)* stats of discrete spot rate models.

		M(1,1)	M(1,2)	M(2,1)	M(2,2)	M(3,3)	M(4,4)
Discretized Single Factor Diffusion Models	RW	59.52	18.12	12.74	225.09	73.57	194.12
	Lognormal	61.74	0.78	3.64	151.37	75.64	131.92
	Dothan	61.92	2.45	3.62	147.53	74.32	125.84
	PCEV	58.57	12.70	8.99	196.60	71.43	171.49
	Vasicek	55.22	15.66	104.21	181.86	24.33	174.12
	CIR	94.09	5.69	47.93	122.04	22.34	124.55
	CKLS	81.02	9.13	69.41	143.97	22.07	146.92
	Nonlinear drift	44.12	3.79	54.61	133.48	20.93	132.54
GARCH Models	No drift	58.19	4.45	16.17	13.96	40.88	12.47
	Linear drift	22.41	0.85	12.81	6.75	15.27	4.38
	Nonlinear drift	21.46	2.39	16.32	5.64	12.19	3.61
	No drift CEV	58.43	3.01	10.15	13.66	40.86	12.88
	Linear drift CEV	23.21	0.28	9.73	7.52	17.35	5.45
	Nonlinear drift CEV	21.39	-0.13	10.35	7.25	12.35	6.17
Markov Regime Switching Models	No drift CEV	57.86	2.39	3.39	18.90	60.22	17.45
	Linear drift CEV	31.50	-0.20	-0.35	8.42	28.46	4.76
	Nonlinear drift CEV	26.86	0.89	1.00	13.46	28.24	11.37
	No drift GARCH	51.85	2.45	7.69	3.85	44.86	2.61
	Linear drift GARCH	37.69	1.07	12.61	4.10	35.07	3.17
	Nonlinear drift GARCH	29.15	0.51	12.99	2.85	23.23	1.01
	No drift CEV GARCH	54.93	2.04	2.77	6.67	53.51	7.82
	Linear drift CEV GARCH	31.57	0.18	-0.60	7.72	29.48	4.59
	Nonlinear drift CEV GARCH	24.65	1.67	-0.21	4.64	24.29	3.45
Jump Diffusion Models	No drift CEV	49.93	-0.70	1.57	57.81	42.22	54.03
	Linear drift CEV	24.56	-0.52	2.31	38.89	23.48	42.32
	Nonlinear drift CEV	22.75	-0.40	2.45	40.36	21.01	43.16
	No drift GARCH	51.06	2.08	0.72	2.30	45.26	3.10
	Linear drift GARCH	31.40	-0.41	4.85	2.22	31.03	3.23
	Nonlinear drift GARCH	27.95	0.00	7.63	1.83	23.56	3.15
	No drift CEV GARCH	50.69	2.47	0.75	2.22	45.02	3.00
	Linear drift CEV GARCH	30.99	-0.46	4.66	2.25	30.69	3.19
	Nonlinear drift CEV GARCH	27.09	-0.60	5.28	1.38	23.44	2.74

Notes: (1) The sample period is from July 22, 1996 and August 26, 2004 with 1954 observations; (2) M(1,1) checks the autocorrelations in level, M(1,2) checks the ARCH-in-mean, M(2,1) checks the leverage effects, and M(2,2), M(3,3) and M(4,4) checks autocorrelations in higher order moments; (3) A lag truncation order p = 10 is used; (4) Upper-tailed N(0,1) critical value (e.g., 1.645 at 5% level) is used for specification tests.